

Optimization notes

Ying-Jia Lin

National Cheng Kung University

May 3rd, 2021

Directional derivative

From a starting point \underline{x}_0 and a given direction \underline{u} :

- $\underline{x}(\lambda) = \underline{x}_0 + \lambda \underline{u}$
 - λ is a scalar.
- $d\underline{x} = \underline{u}d\lambda$
 - For a small change in λ .
- $F(\lambda) = f(\underline{x}_0 + \lambda \underline{u})$

$$\begin{aligned}dF &= df = (\nabla f(\underline{x}))^\top d\underline{x} \\&= (\nabla f(\underline{x}))^\top \underline{u}d\lambda = \nabla^\top f \underline{u} \lambda\end{aligned}$$

- $\frac{df}{d\lambda} = \nabla^\top f \underline{u}$
 - If f is minimized at $\underline{x}^* = \underline{x}_0 + \lambda \underline{u}$, then:
 - $\nabla f(\underline{x}^*)^\top \underline{u} = 0$
 - gradient f evaluated at the minimum point is orthogonal to \underline{u} .

Weierstrass Theorem

If $f(\underline{x})$ is continuous on a nonempty feasible set that is closed and bounded, then $f(\underline{x})$ has a global minimum in this set.

- ▶ A set S is bounded if for any point \underline{x} in S , we have $\underline{x}^T \underline{x} < c$
 - ▶ c is a finite positive number.

Single-variable unconstrained optimization

- Necessary condition
 - If a function $f(x)$ has a local minimum at $x = x^*$, and $f'(x)$ exists as a finite number at $x = x^*$, then $f'(x^*) = 0$.
 - x^* at $f'(x^*) = 0$ is called stationary point.
- Sufficient condition
 - Suppose $f'(x^*) = f''(x^*) = \dots = f^{(m-1)}(x^*) = 0$, but $f^{(m)}(x^*) \neq 0$, then $f(x^*)$ is:
 - 1. a local minimum if $f^{(m-1)}(x^*) > 0$ and m is even.
 - 2. a local maximum if $f^{(m-1)}(x^*) < 0$ and m is even.
 - 3. neither a maximum nor a minimum if m is odd.

Multi-variable unconstrained optimization (1)

Definition of r^{th} differential of function f :

$$d^r f(\underline{x}^*) = \sum_{i=1}^n \sum_{j=1}^n \cdots \sum_{k=1}^n h_i h_j \cdots h_k \frac{\partial^r f(\underline{x}^*)}{\partial x_i \partial x_j \cdots \partial x_k}$$

Example

When (order) $r = 2$ and (number of variables) $n = 3$, we have:

$$\begin{aligned} d^2 f(\underline{x}^*) &= d^2 f(x_1^*, x_2^*, x_3^*) = \sum_{i=1}^3 \sum_{j=1}^3 h_i h_j \frac{\partial^2 f(\underline{x}^*)}{\partial x_i \partial x_j} \\ &= h_1^2 \frac{\partial^2 f(\underline{x}^*)}{\partial x_1^2} + h_2^2 \frac{\partial^2 f(\underline{x}^*)}{\partial x_2^2} + h_3^2 \frac{\partial^2 f(\underline{x}^*)}{\partial x_3^2} \\ &\quad + 2h_1 h_2 \frac{\partial^2 f(\underline{x}^*)}{\partial x_1 \partial x_2} + 2h_2 h_3 \frac{\partial^2 f(\underline{x}^*)}{\partial x_2 \partial x_3} + 2h_1 h_3 \frac{\partial^2 f(\underline{x}^*)}{\partial x_1 \partial x_3} \end{aligned}$$

Multi-variable unconstrained optimization (2)

- Necessary condition

$$\frac{\partial f(\underline{x}^*)}{\partial x_1} = \frac{\partial f(\underline{x}^*)}{\partial x_2} = \dots = \frac{\partial f(\underline{x}^*)}{\partial x_n} = 0$$

- In vector form, $\nabla f(\underline{x}^*) = 0$.
- \underline{x}^* at $\nabla f(\underline{x}^*) = 0$ is called stationary point.
- Sufficient condition
 - For a stationary point at $\underline{x} = \underline{x}^*$:
 - if the Hessian matrix of $f(\underline{x})$ evaluated at $\underline{x} = \underline{x}^*$ is **positive definite**, then \underline{x}^* is a local minimum.
 - if the Hessian matrix of $f(\underline{x})$ evaluated at $\underline{x} = \underline{x}^*$ is **negative definite**, then \underline{x}^* is a local maximum.

Hessian Matrix

- $(H_f)_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_n}$

- In matrix form: $H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$

Definiteness

- A matrix is positive definite if all its **eigenvalues** are positive.
 - If some of the eigenvalues are positive and some are zero, then the matrix is positive semidefinite.
- **Checking the sign of the determinants** is an alternative way to determine the definiteness of a matrix.
- A twice differentiable function is convex if and only if its Hessian matrix is positive semi-definite.
 - The function is strictly convex if the Hessian matrix is positive definite.

Multivariable optimization with equality constraints

Lagrange Multiplier Theorem

- Suppose the point \underline{x}^* minimizes $f(\underline{x})$ and satisfies the equality constraints: $h_j(\underline{x}^*) = 0$, for $j = 1, 2, \dots, m$
- Assume that the constraint gradients $\nabla h_j(\underline{x}^*)$ are linealy independent.
- Then there exists a unique set λ_j^* ($j = 1, 2, \dots, m$) satisfying:

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^m \lambda_j^* \frac{\partial h_j}{\partial x_i} = 0$$

where $i = 1, 2, \dots, n$.

Simplex method: Introduction

Definition of Simplex

The geometric figure formed by a set of $n + 1$ points in an n -dimensional space is called a simplex.

- When the points are equidistant, the simplex is said to be *regular*.
- In two dimensions the simplex is a triangle, and in three dimensions, it is a tetrahedron.
- The simplex method was originally given by Spendley et al. and was developed later by Nelder and Mead.

Simplex method: Steps

Choose a reflection coefficient $\alpha > 0$, an expansion coefficient $\gamma > 1$, and a contraction coefficient $0 < \beta < 1$.

1. Identify \underline{x}_{\min} and \underline{x}_{\max} among $\{\underline{x}_1, \underline{x}_2, \dots, \underline{x}_{n+1}\}$
 - such that $f(\underline{x}_{\min})$ is the minimum and $f(\underline{x}_{\max})$ is the maximum of all the $f(\underline{x}_i)$, for $i = 1, 2, \dots, n + 1$.
 - If $|\underline{x}_{\max} - \underline{x}_{\min}| < \epsilon$, stop. The minimum is at \underline{x}_{\min} .
 - Otherwise, let \underline{x}_a be the averaged position of $\{\underline{x}_1, \underline{x}_2, \dots, \underline{x}_{n+1}\}$, **excluding** \underline{x}_{\max} , and go to step 2.
2. Let the reflection point $\underline{x}_r = \underline{x}_a + \alpha(\underline{x}_a - \underline{x}_{\max})$.
 - If $f(\underline{x}_{\min}) > f(\underline{x}_r)$, let the expansion point $\underline{x}_e = \underline{x}_a + \gamma(\underline{x}_r - \underline{x}_a)$, and go to step 3.
 - Otherwise, go to step 4.

Simplex method: Steps

3. If $f(\underline{x}_r) > f(\underline{x}_e)$, the point \underline{x}_{\max} is **replaced** by \underline{x}_e .
 - Otherwise, \underline{x}_{\max} is **replaced** by \underline{x}_r . A new set of $n+1$ points is formed. Then go to step 1.
4. If the second largest $f(\underline{x}_i) > f(\underline{x}_r)$, then \underline{x}_{\max} is **replaced** by \underline{x}_r to form a new set of $n+1$ points, and go to step 1.
 - Otherwise, go to step 5.
5. Let \underline{x}_p be defined such that $f(\underline{x}_p) = \min \{f(\underline{x}_r), f(\underline{x}_{\max})\}$ and let the contraction point $\underline{x}_c = \underline{x}_a + \beta(\underline{x}_p - \underline{x}_a)$.
 - If $f(\underline{x}_c) > f(\underline{x}_p)$, **replace** \underline{x}_j by $\underline{x}_j + (\underline{x}_{\min} - \underline{x}_j)/2$, for $j = 1, 2, \dots, n+1$, and go to step 1.
 - Otherwise, \underline{x}_c **replaces** \underline{x}_{\max} to form a new set of $n+1$ points, and go to step 1.

Transformation techniques

It may be possible to convert a constrained optimization problem into an unconstrained one by making a change of variables.

If lower and upper bounds on x_i are specified as:

$$l_i \leq x_i \leq u_i$$

which can be satisfied by transforming the variable x_i as:

$$x_i = l_i + (u_i - l_i) \sin^2 y_i$$

where y_i is the new variable, which can take any value.

- If x_i is restricted to lie in the interval $(0, 1)$, $x_i = \sin^2 y$

Transformation techniques

- ① The constraints $g_i(X)$ must be very simple.
- ② For certain constraints, it may not be possible to find the necessary transformation.
- ③ If it is not possible to eliminate all the constraints by making a change of variables.
 - It may be better not to use the transformation at all.
 - However, the partial transformation may sometimes produce a distorted objective function which might be harder to minimize than the original function.

Penalty function method

Find X which minimizes $f(X)$ is converted into an unconstrained minimization problem by constructing a function of the form:

$$\phi_k = \phi(X, r_k) = f(X) + r_k \sum_{j=1}^m G_j[g_j(X)]$$

- g_j : inequality constraints
- G_j : some function of the constraint g_j
- r_k : penalty parameter (a positive constant)
- $r_k \sum_{j=1}^m G_j[g_j(X)]$ is the penalty term.